1

PARAMETER ESTIMATION IN ENGINEERING AND SCIENCE

JAMES V. BECK
Department of Mechanical Engineering and
Division of Engineering Research
Michigan State University

and

KENNETH J. ARNOLD
Department of Statistics and Probability
Michigan State University

John Wiley & Sons
New York London Sydney Toronto

2

EXTENSION SERIES

r the Engineer, First Series
r the Engineer, Second Series
Systems for Missiles and Space

ials at Elevated Temperatures
es
ons of Future Electronics
Technology
acecraft
Engineering
gineer, First Series
hysics for the Engineer,

e Engineer and Scientist

he Utilization of Solar Energy

SECOND SERIES

MODERN MATHEMATICS FOR THE ENGINEER

ARTHUR ERDÉLYI
BERNARD FRIEDMAN
JOHN W. MILES
RALPH S. PHILLIPS
J. BARKLEY ROSSER
WILLIAM FELLER
DAVID BLACKWELL
RICHARD BELLMAN
GEORGE B. DANTZIG
SAMUEL KARLIN
STANISLAW M. ULAM
RAYMOND REDHEFFER
SUBRAHMANYAN CHANDRASEKHAR
PAUL R. GARABEDIAN
DAVID YOUNG
GEORGE PÓLYA

Edited by
EDWIN F. BECKENBACH
Professor of Mathematics
University of California
Los Angeles

With an Introduction by
MAGNUS R. HESTENES

McGRAW-HILL BOOK COMPANY, INC. 1961
NEW YORK TORONTO LONDON

PROBABILISTIC SYSTEMS ANALYSIS

AN INTRODUCTION TO PROBABILISTIC MODELS, DECISIONS, AND APPLICATIONS OF RANDOM PROCESSES

ARTHUR M. BREIPOHL

PROBABILISTIC SYSTEMS ANALYSIS

AN INTRODUCTION TO PROBABILISTIC MODELS, DECISIONS, AND APPLICATIONS OF RANDOM PROCESSES

ARTHUR M. BREIPOHL
Oklahoma State University

JOHN WILEY & SONS, INC.
NEW YORK · LONDON · SYDNEY · TORONTO

Thus independence is not required for the first result. Now

$$E[(Y-\mu_Y)^2] \simeq E\left\{\left[\sum_{i=1}^{n} a_i(X_i-\mu_i)\right]^2\right\}$$

$$\simeq \sum_{j=1}^{n}\sum_{i=1}^{n} a_i a_j E[(X_i-\mu_i)(X_j-\mu_j)].$$

$$\sigma_Y^2 \simeq \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + \sum_{\substack{j=1 \\ i\neq j}}^{n}\sum_{i=1}^{n} a_i a_j \sigma_{ij} \qquad (6\text{-}18)$$

where $\qquad \sigma_{ij} = E[(X_i-\mu_i)(X_j-\mu_j)] = E[X_iX_j] - \mu_i\mu_j.$

6-8 SYNTHETIC SAMPLING (MONTE CARLO TECHNIQUE)

The methods discussed so far to find the distribution of

$$Y = g(X_1, X_2, \ldots, X_n)$$

have been approximations or too involved to be practical for large problems. In this section a very simple and intuitively satisfying method is presented. Its only drawbacks are that it requires a digital computer and general parametric results are not obtained, thus limiting the applicability to synthesis.

It is assumed that $Y = g(X_1, \ldots, X_n)$ is known and that the joint density $f_{X_1,X_2,\ldots,X_n}$ is known. Now if a sample value of each random variable were known (say $X_1 = x_{11}, X_2 = x_{12}, \ldots, X_n = x_{1n}$), then a sample value of Y could be computed [say $y_1 = g(x_{11}, x_{12}, \ldots, x_{1n})$]. Then if another set of sample values were chosen for the random variables (say $X_1 = x_{21}, \ldots, X_n = x_{2n}$), then $y_2 = g(x_{21}, x_{22}, \ldots, x_{2n})$ could be computed.

If one had the time one could compute many such sample values of Y. The computer actually supplies the speed that makes many such calculations possible. There is just one problem. How does the computer select the different values of $X_1, X_2, \ldots, X_n$?

If each of the random variables had a uniform distribution between 0 and 1, numbers for each random variable could be chosen from a table of random numbers. Actually, computer routines generate pseudorandom numbers which may be used.

Consider the following case. Let the random variables $X_1, X_2, \ldots, X_{30}$ be independent and each uniformly distributed between zero and one, and let $Y = g(X_1, X_2, \ldots, X_{30})$ be a known function. Then the computer program

to compute an
ing basic steps

no

If a plot i
histogram. T
30 mutually e
of samples t
histograms.)
The case
have a distrib
ing procedure
distributed b
the random s
For examp
Then

This is show
Notice F
sample of U,

6
ow
$_j)]$.
(6-18)
μ_j.
TECHNIQUE)

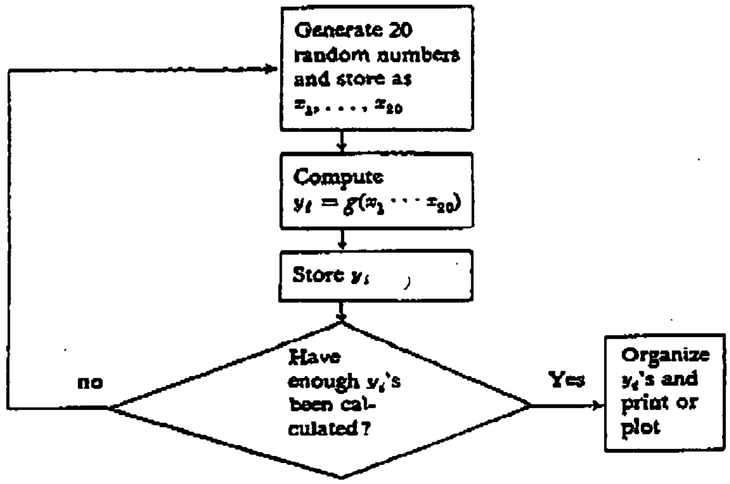
or large problems.
thod is presented.
and general para-
ility to synthesis.
at the joint density
om variable were
ample value of Y
if another set of
$= x_{i1}, \ldots, X_n =$

mple values of Y.
such calculations
mputer select the

bution between 0
n from a table of
te pseudorandom

$X_1, X_2, \ldots, X_{20}$
ero and one, and
computer program

to compute an approximation to the distribution of Y consists of the following basic steps.



If a plot is desired it may be convenient to plot a usual bar chart or histogram. This method simply calls for breaking the range of Y into say 30 mutually exclusive cells of the same size and plotting vertically the number of samples that fell into that cell. (See Chapter VIII for more details of histograms.)

The case of uniformly distributed variables was considered. Now let X_i have a distribution function F_{X_i}. To obtain a random sample of X_i the following procedure may be used. Select a random sample of U which is uniformly distributed between 0 and 1. Call this random sample u_1. Then $F_X^{-1}(u_1)$ is the random sample of X (see Example 6-7).

For example, suppose that X is uniformly distributed between 10 and 20. Then

$$
\begin{aligned}
F_{X_i}(x) &= 0, && x < 10,\\
&= (x - 10)/10, && 10 \leq x < 20,\\
&= 1, && x \geq 20.
\end{aligned}
$$

This is shown in Figure 6-29.

Notice $F_X^{-1}(u) = 10u + 10$. Thus if the value .250 were the random sample of U, then the corresponding random sample of X would be 12.5.

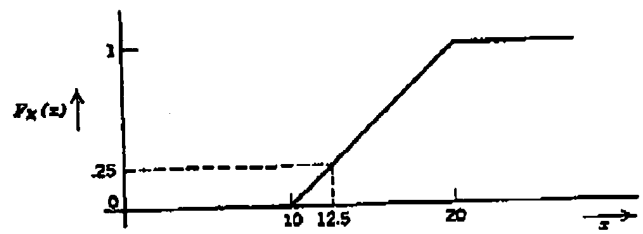


Figure 6-29

As another example suppose that X is normally distributed with mean 0 and variance 1. Then a random value of .250 for U would correspond to $-.67$ for a value of X. This result follows from a table of the normal distribution function. Practically, most computers automatically generate standard normal random variables. If this is the case then random sample of a normal random variable Y with mean μ and a variance σ^2 may be generated from a standard normal random variable X by recalling that

$$X = \frac{Y - \mu}{\sigma}$$

or

$$Y = X\sigma + \mu. \tag{6-19}$$

Equation 6-19 can be easily checked by finding the characteristic function of Y.

The only difference when the random variables are dependent is that the dependence must be taken into account when the random samples are generated. We assume that the dependence is expressed in terms of conditional distributions. If this is not the case the joint distributions can always be reduced to the required conditional distributions.

For illustration consider three dependent random variables: X_1, X_2, and X_3. We first generate a random sample of X_1 by the same methods discussed above. Call this sample x_{11}. We next generate a random sample of X_2 using $F_{X_2|X_1=x_{11}}$, by the same method used before. Call this sample x_{12}. Then we use $F_{X_3|X_1=x_{11}, X_2=x_{12}}$ to generate x_{13}. Thus nothing changes except that the conditional distribution functions are used in generating the random samples.

With a fast digital computer thousands of simulations can be run in reasonable times. Monte Carlo solutions often involve 10,000 or more simulations. An example is given in the next chapter.

6-9 SUMM

The purpose
problem of
distribution
First the
solved by (6-
Next Y =
of independe
density funct
where Y is a
generalizatio
The gene
of solution
example, an
Two app
Taylor series
used. Then a
Referenc

6-10 PRO

1. X has a n

Find the

2. The pow
random

Find the

3. The outp
when X

4. The outp

Find the

6-9 SUMMARY

The purpose of this chapter was to consider the important engineering problem of finding the distribution of $Y = g(X_1, X_2, \ldots, X_n)$ where the distribution of the X_i's is known.

First the problem of $Y = g(X)$ was considered and this problem was solved by (6-2). Examples were given to illustrate its application.

Next $Y = \sum_{i=1}^{n} X_i$ was considered, and it was shown that in the case of independent random variables the solution involved convolution of the density functions or multiplication of the characteristic functions. The case where Y is a linear combination of the X_i's was shown to be only a slight generalization of this problem.

The general problem was then considered and although a general method of solution was outlined, the difficulty of solution was illustrated by an example, and approximations were suggested.

Two approximations for the general solution were described. First a Taylor series approximation, moments, and the central limit theorem were used. Then a Monte Carlo method was suggested.

References B1, D3, D4, and P2 provide additional reading.

6-10 PROBLEMS

1. X has a normal density function with mean 1 and variance 2.

$$Y = \tfrac{1}{2}X - 1.$$

Find the density of Y.

2. The power P dissipated in a resistor is $P = I^2R$. Assume $R = 2$ and I is a random variable with a normal density.

$$f_I(i) = \frac{1}{\sqrt{2\pi}} e^{\frac{-i^2}{2}}.$$

Find the density function of P.

3. The output of a full wave rectifier is $Y = |X|$. Find the density function of Y when X has a uniform density from -1 to $+1$.
4. The output of a square law detector is

$$Y = aX^2, \qquad a > 0.$$

Find the density function of Y in terms of the density function f_X of X.

sistors are pur-

(7-7)

hown in Figure
iformly distrib-

y between 1/3.5
re not absolute
lescribed by the
d the remainder
nsity function is
arger than 1/3.5
rd deviation can
nce limits are set
usually occur in

ne the standard
elation

is usually a good assumption. When the standard deviation of the parts are known, then the mean and variance of the output can be computed using the approximation developed in the last chapter.

Again something must be assumed to describe the output distribution and the probability of being out of tolerance. It is suggested that if enough variables are involved and the function is approximately a linear combination, then a normal density may be assumed. Then the probability of being outside tolerance limits can be computed.

We now show an example of a tolerance problem which is solved using a Monte Carlo approach.

EXAMPLE 7-2

This example, a simplification of a problem that actually occurred in practice, was worked by the method shown, and the results actually obtained in manufacture corresponded with the theoretical results.

The simplified version is shown in Figure 7-4. The bar will fit within the bracket if $Y_a < Y_b$ and there will be interference (or no fit) if $Y_b < Y_a$. In the actual case Y_a and Y_b involved 41 dimensions and the configuration was more complicated than simply a sum of lengths. Actually the equations for Y_a and Y_b involved arcs and angles, thus various trigonometric functions were involved.

The problem was solved by finding the probability distribution of $Z = Y_b - Y_a$ by Monte Carlo sampling. Note that if $Z > 0$ there is no problem, while if $Z < 0$ there will be no fit and the parts cannot be assembled.

The problem arose because just as production and assembly were about to start it was discovered that interference was possible. (As customary, the intent was to tighten the tolerances on each part until no interference is possible at the worst case, but the designer made a mistake in his worst case calculations.) Then the question was, what is the probability of interference? If it is low enough then it would be better to have a few that would not fit, rather than wait and spend the extra money to redesign some of the parts.

To find the probability density of Z via a Monte Carlo technique, one must have $Z = g(X_1, \ldots, X_{41})$ and must know the joint distribution of

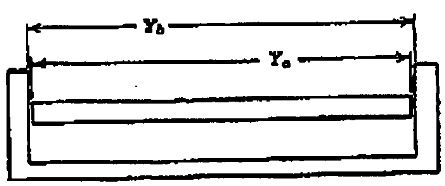


Figure 7-4

$X_1, \ldots, X_{41}$. The function g was found from the drawings using trigonometry. A part of the equation is shown in Figure 7-5 simply to illustrate the

$$
\begin{aligned}
Y = (R + W)\sin\Bigg\{\theta - \Bigg[\sin^{-1}\left(\frac{H_{21}}{R}\right) &+ 2\sin^{-1} \\
&\times\left(\frac{\sqrt{\left(\sqrt{r^2 - H_{11}^2} - \sqrt{r^2 - (S-V)^2}\right)^2 + [(S-V) - H_{11}]^2} - \dfrac{H_{21} - H_{12}}{2}}{2(R+W)}\right)\Bigg]\Bigg\} \\
&+ \sqrt{(S-V)^2 + \left(Z + r - \sqrt{r^2 - (S-V)^2}\right)^2} \\
&\cdot \sin^{-1}\Bigg\{\tan^{-1}\left(\frac{Z + r - \sqrt{r^2 - (S-V)^2}}{(S-V)}\right) + \tan^{-1}\left(\frac{\sqrt{r^2 - (S-V)^2} - \sqrt{r^2 - V^2}}{S}\right) \\
&- \Bigg[180^\circ - \Bigg[\Bigg[\frac{\left(180^\circ - 2\sin^{-1}\dfrac{\sqrt{S^2 + \left(\sqrt{r^2 - (S-V)^2} - \sqrt{r^2 - V^2}\right)^2}}{2(R+W)}\right)}{2}\Bigg] \\
&+ \Bigg[90^\circ - \Bigg[\theta - \left(\sin^{-1}\left(\frac{H_{21}}{R+W}\right) + 2\sin^{-1}\right. \\
&\times\left.\left(\frac{\sqrt{\left(\sqrt{r^2 - H_{11}^2} - \sqrt{r^2 - (S-V)^2}\right)^2 + [(S-V) - H_{11}]^2} - \dfrac{H_{21} - H_{12}}{2}}{2(R+W)}\right)\right)\Bigg]\Bigg]\Bigg]\Bigg]\Bigg\}
\end{aligned}
$$

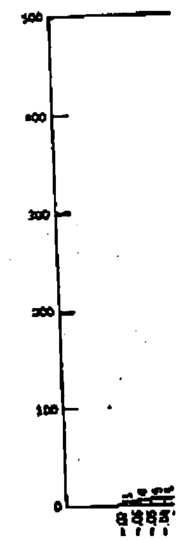
Figure 7-5

form. The various dimensions were assumed to be independent and equally likely between their upper and lower tolerance limits. The result of 8000 simulations is shown in Figure 7-6.

Note that the results appear nearly normal and that interference occurred 71 times in 8000 simulations. Based on these results it was decided to produce units without a design change and to rebuild those few on which interference did occur. The results of the actual assembly operation corresponded very well with the prediction that 71/8000 would not fit.

Summary

Tolerances of parameters can be combined by Monte Carlo methods or by Taylor series approximation. Both produce the probability of being outside certain limits and are less conservative than deterministic tolerance studies. Note that the analysis described above is the basis for deciding which tolerances to assign to what parts.



7-3 CLOS

Consider th
logic, solid
networks of
We first
basic mathe
Conside
closure time
switch and
There w
are closed o

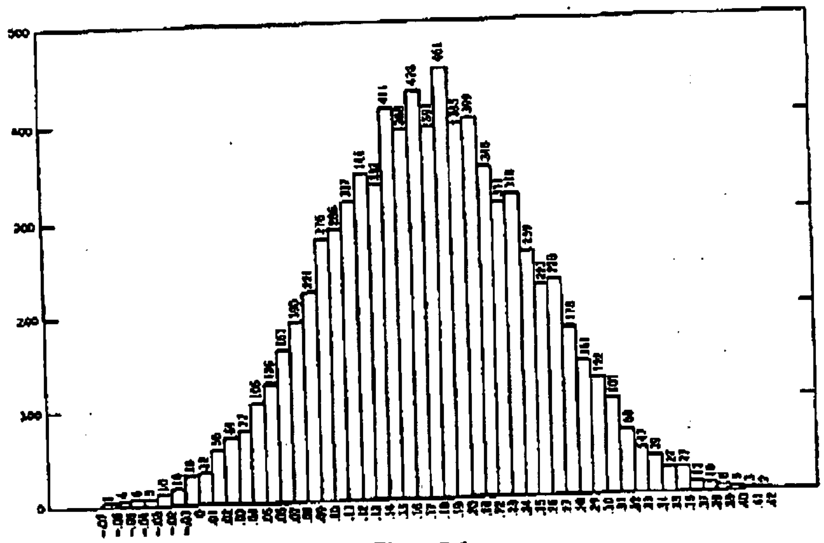


Figure 7-6

7-3 CLOSURE TIME OF SWITCHING CIRCUITS

Consider that the distribution of closure times of switches (e.g., relays, fluid logic, solid state devices) is known. We wish to know the closure time of networks of these switches.

We first model the circuit with a mathematical function and then solve the basic mathematical problems in more general terms (so we can use it later).

Consider two switches in series as shown in Figure 7-7. What is the closure time T of the series circuit in terms of T_1, closure time of the first switch and T_2, closure time of the second switch?

There will be a closed circuit from a to b at the time when both switches are closed or when the latter of the two switches closes. That is

$$T = \max(T_1, T_2).$$

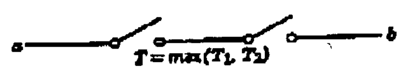


Figure 7-7

median value of the conditional distribution of β given Y, $f(\beta|Y)$. If the density $f(\beta|Y)$ in addition to being symmetric is also unimodal, the mean, median, and mode will all be at the *same* location. Hence when $f(\beta|Y)$ is symmetric about the parameter vector β and is also unimodal, b_{SEL} is b_{MAP}. When the distribution is *not* symmetric or *not* unimodal b_{SEL} and b_{MAP} are rarely the same. Some nonsymmetric unimodal probability densities are depicted in Fig. 4.2. Note that the modes do not coincide with the means. This causes the parameters b_{SEL} given by (4.7.3) and associated with the *mean* to be *not* equivalent to those given by the *mode* which are indicated by (4.7.2).

The conditional probability density $f(\beta|Y)$ used in (4.7.2) can be written in terms of other densities using the form of Bayes's theorem written as

$$f(\beta|Y)=\frac{f(Y|\beta)f(\beta)}{f(Y)} \qquad (4.7.4)$$

The probability density $f(\beta)$ contains the prior information known regarding the parameter vector β. Notice that the parameters appear only in the numerator of the right side of (4.7.4); this numerator can also be written as

$$f(Y,\beta)=f(Y|\beta)f(\beta) \qquad (4.7.5)$$

Then the necessary conditions given by (4.7.2) can be written equivalently as

$$\left.\frac{\partial \ln[f(Y,\beta)]}{\partial\beta_i}\right|_{b_{MAP}}=\left.\frac{\partial \ln[f(Y|\beta)]}{\partial\beta_i}\right|_{b_{MAP}}+\left.\frac{\partial \ln[f(\beta)]}{\partial\beta_i}\right|_{b_{MAP}}=0 \qquad (4.7.6)$$

since the maximum of $f(Y,\beta)$ exists at the same location as the maximum of its natural logarithm.

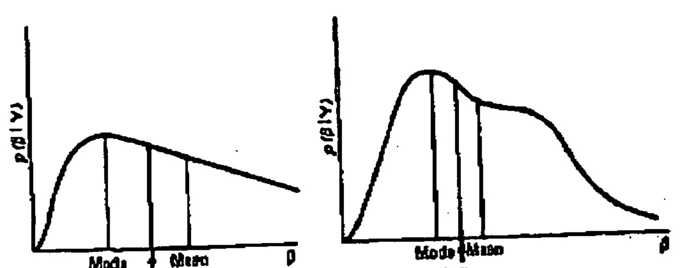


Figure 4.2 Some nonsymmetric conditional probability densities.

The estimators b_{MAP}, b_{SEL} are described without reference to the linearity or nonlinearity of the expected value of Y in the β's nor to the independence of the Y_i's. Under some assumptions about the structure of η_i and under some assumptions about the prior distribution of the β's, the MAP and SEL procedures are equivalent in arithmetic to certain least squares or Gauss–Markov procedures.

4.8 COST

Methods of collecting data and analyzing them must be coordinated. If observations are expensive, sophisticated methods of analysis to extract all pertinent information are justified. Sometimes more expensive methods of collecting data yield net returns by drastically reducing the cost of analysis. Increased costs due to collecting more data or using more sophisticated methods of analysis may or may not reduce the cost occasioned by the degree to which the estimate is incorrect. Some remarks in Chapter 3 were directed to these matters.

4.9 MONTE CARLO METHODS

One method for investigating the effects of nonlinearity or various other effects that are difficult to analyze otherwise is called the Monte Carlo method. Actually, what we describe is sometimes referred to as the "crude" Monte Carlo method. More sophisticated Monte Carlo methods often provide the same amount of information as the crude method but at a lower cost [1].

The Monte Carlo method can be used to investigate analytically the properties of a proposed estimation method. To simulate a series of experiments on the computer we proceed as follows:

1. Define the system by prescribing (*a*) the model equation, also called regression function, (*b*) the way in which "errors" are incorporated in the model of the observations, (*c*) the probability distribution of all the errors and, where applicable, (*d*) a prior distribution. Assign "true" values to all the parameters (β) in the regression function and to those in the distribution of error.
2. Select a set of values of the independent variables. Then calculate the associated set of "true" values of η from the regression equations.
3. Use the computer to produce a set of errors ϵ drawn from the prescribed probability distribution. For most computers programs are

available which can generate a stream of numbers that have all the important characteristics of successive independent observations on a population uniform over the interval (0, 1). Since they are generated by a deterministic scheme, they are not actually random. Such numbers are called *pseudorandom numbers*. Suitable transformations are used to obtain samples for any other distribution.

To obtain a sequence of pseudorandom observations on a normal population with expected value 0 and variance 1, we can make use of the Box–Muller transformation [2]. If u_{2i-1} and u_{2i} are independent (0, 1) random numbers,

$$x_{2i-1}=(-2\ln u_{2i-1})^{1/2}\cos(2\pi u_{2i}) \qquad (4.9.1a)$$

and

$$x_{2i}=(-2\ln u_{2i-1})^{1/2}\sin(2\pi u_{2i}) \qquad (4.9.1b)$$

are independent random observations on a normal distribution with expected value 0 and variance 1. The normal random numbers are then adjusted to have the desired variances and covariances.

The simulated measurements are obtained by combining the errors with the regression values. For additive errors, the ith error is simply added to the ith η value. This then provides simulated measurements.

4. Acting as though the parameters are unknown, we estimate the parameters, denoting the estimates β^*.
5. Replicate the series of simulated experiments N times by repeating steps 3 and 4, each time with a new set of errors.
6. We use appropriate methods to estimate properties of the distribution of parameter estimates. (We consider the estimates actually obtained by our pseudorandom number scheme to be a random sample from the distribution of all possible estimates.) The expected value of our parameter estimator is estimated by the mean of our parameter estimates,

$$\bar{\beta}_j^*=\frac{1}{N}\sum_{i=1}^{N}\beta_{ji}^* \qquad (4.9.2)$$

where β_{ji}^* is the jth component of the β^* found on the ith replication. If β^* may be a biased estimator, $\beta^*-\beta$ is an estimate of the bias. If it is not clear whether or not β^* is biased the size of $\beta^*-\beta$ needs to be compared with an estimate of its variance-covariance matrix.

The variances and covariances of the distribution of β^* may be

estimated by

$$\text{est. cov}(\beta_j^*,\beta_k^*)=\frac{1}{N-1}\sum_{i=1}^{N}\left(\beta_{ji}^*-\bar{\beta}_j^*\right)\left(\beta_{ki}^*-\bar{\beta}_k^*\right) \qquad (4.9.3a)$$

If β^* is known to be unbiased, we can make use of our knowledge of β and use a slightly more efficient estimator

$$\text{est. cov}(\beta_j^*,\beta_k^*)=\frac{1}{N}\sum_{i=1}^{N}(\beta_{ji}^*-\beta_j)(\beta_{ki}^*-\beta_k) \qquad (4.9.3b)$$

If β^* is biased, the right side of (4.9.3b) which are estimates of mean square error and corresponding product moments, may be more interesting than variances and covariances. If we use actual experiments rather than simulated ones (4.9.3b) will be not available although (4.9.2) and (4.9.3a) are.

The flexibility of the above simulation procedure is great. We can estimate the sample properties for any model, linear or nonlinear, and for any parameter values. We can estimate the effect of different probability distributions upon ordinary least squares estimation or other estimation methods. Many other possibilities also exist. An example of a Monte Carlo simulation is given below and another one is given in Section 6.9. These simulations can be accomplished on a modern high-speed computer at a small fraction of the cost, in time and money, of a comparable set of physical experiments.

The great power of the Monte Carlo procedure is that we can investigate the properties of estimators in cases for which the character of the estimators cannot be derived. To demonstrate the validity of a Monte Carlo procedure an example is considered which is simple enough to be analyzed without recourse to simulation. We investigate estimating β in the model $\eta_i=\beta X_i$ for the case of additive, zero mean, constant variance, uncorrelated errors; that is

$$y_i=\eta_i+\epsilon_i, \quad E(\epsilon_i)=0, \quad V(\epsilon_i)=\sigma^2, \quad E(\epsilon_i\epsilon_j)=0 \quad \text{for } i\neq j$$

The distribution of ϵ_i is uniform in the interval $(-.5,.5)$; each ϵ_i is found using a pseudorandom number generator. There are no errors in X_i and there is no prior information.

The X_i values are $X_i=i$ for $i=1,2,\ldots,10$ and $\beta=1$. For the kth set of simulated measurements, β_k^* is found using the ordinary least squares

estimator,

$$\beta_k^* = \left[\sum_{i=1}^{10} X_i Y_{i,k}\right]\left[\sum_{i=1}^{10} X_i^2\right]^{-1}$$

The estimated expected value of β_k^*, (4.9.2), the estimated variance of β_k^*, (4.9.3a), and the estimated mean square error of β_k^*, (4.9.3b), are obtained by using

$$\bar{\beta}^* = \frac{1}{10}\sum_{k=1}^{10} \beta_k^*, \text{ est. } V(\beta^*) = \frac{1}{9}\sum_{k=1}^{10}\left(\beta_k^* - \bar{\beta}^*\right)^2$$

$$\text{est. mean square error } (\beta^*) = \frac{1}{10}\sum_{k=1}^{10}(\beta_k^* - 1)^2$$

For independent sets of errors, estimates were calculated for $N = 5, 25, 50, 100, 200$, and 500. The results are shown in Table 4.1 where the estimated standard deviation and estimated root mean square error are given rather than their squares. In Table 4.2 comparable results for a simulation involving normal errors are given. The variance of ϵ_i in this case was taken as 1/12, the same as the variance for the uniform case.

In both Tables 4.1 and 4.2 the sample mean $\bar{\beta}^*$ tends to approach the ture value of 1 as N becomes large. Hence β^* is an unbiased estimator of β. Also the estimated standard error of β^* and estimated root mean square error tend to their common exact value

$$\left\{\sigma^2\left[\sum X_i^2\right]^{-1}\right\}^{1/2} = \left\{\frac{1/12}{385}\right\}^{1/2} = 0.014712$$

Table 4.1 Monte Carlo Simulation for $\eta_i = \beta X_i$, with $\beta = 1$ and $X_i = i$, $i = 1, 2, \ldots, 10$. Uniform Distribution of Errors

Sample Size	$\bar{\beta}^*$	Est. Std Dev (β^*)	Est. Root Mean Square Error (β^*)
5	1.0044	0.00950	0.00958
25	1.0014	0.01616	0.01589
50	0.9992	0.01350	0.01339
100	0.9996	0.01425	0.01418
200	1.0018	0.01440	0.01448
500	0.9987	0.01415	0.01419

Table 4.2 Monte Carlo Simulation for $\eta_i = X_i$, with $\beta = 1$ and $X_i = i$, $i = 1, 2, \ldots, 10$. Normal Distribution of Errors

Sample Size	$\bar{\beta}^*$	Est. Std Dev (β^*)	Est. Root Mean Square Error (β^*)
5	1.0021	0.01156	0.01055
25	0.9969	0.01608	0.01606
50	0.9972	0.01496	0.01507
100	0.9973	0.01486	0.01502
200	0.9995	0.01410	0.01407
500	0.9997	0.01480	0.01478

This example shows that the number of simulations N must be quite larg in order to provide accurate estimates of the variance of the paramete estimate. Such simulations are still inexpensive compared to actual experi ments to determine the variance. Moreover, methods are available fo making the simulation procedure more efficient [1].

REFERENCES

1. Hammersley, J. M. and Handscomb, D. C., *Monte Carlo Methods*, Methuen & Co. Ltd London, 1964.
2. Box, G. E. P. and Muller, M. E., "A Note on the Generation of Random Norm Deviates," *Ann. Math. Stat.*, 29 (1958), 610–611.